The Patent Office
Concept House
Cardiff Road
Newport
South Wales
NP10 8QQ

I, the undersigned, being an officer duly authorised in accordance with Section 74(1) and (4) of the Deregulation & Contracting Out Act 1994, to sign and issue certificates on behalf of the Comptroller-General, hereby certify that annexed hereto is a true copy of the documents as originally filed in connection with the patent application identified therein.

In accordance with the Patents (Companies Re-registration) Rules 1982, if a company named in this certificate and any accompanying documents has re-registered under the Companies Act 1980 with the same name as that with which it was registered immediately before re-registration save for the substitution as, or inclusion as, the last part of the name of the words "public limited company" or their equivalents in Welsh, references to the name of the company in this certificate and any accompanying documents shall be treated as references to the name with which it is so re-registered.

In accordance with the rules, the words "public limited company" may be replaced by p.l.c., plc, P.L.C. or PLC.

Re-registration under the Companies Act does not constitute a new legal entity but merely subjects the company to certain additional company law rules.

Signed

Dated 6 May 2004

PRIORITY DOCUMENT
SUBMITTED OR TRANSMITTED IN
COMPLIANCE WITH
RULE 17.1(a) OR (b)

Patents Form 1/77

Patents Act 1977
(Rule 16)

The Patent Office

17APR03 E808395 002B01 P01/7700 0.00-0308851.5

THE PATENT OFFICE
16 APR 2003
RECEIVED BY HAND

The Patent Office

Cardiff Road
Newport
South Wales
NP9 1RH

Request for grant of a patent

(See the notes on the back of this form. You can also get an explanatory leaflet from the Patent Office to help you fill in this form)

1. Your reference

JWJ01009GB

2. Patent application number
(The Patent Office will fill in this part)

0308851.5

16 APR 2003

3. Full name, address and postcode of the or of each applicant *(underline all surnames)*

LingVitae AS
Trimveien 6, A562
0372 Oslo
Norway

Patents ADP number *(if you know it)*

86125820001

If the applicant is a corporate body, give the country/state of its incorporation

Norway

4. Title of the invention

Method

5. Name of your agent *(if you have one)*

Gill Jennings & Every

"Address for service" in the United Kingdom to which all correspondence should be sent *(including the postcode)*

Broadgate House
7 Eldon Street
London
EC2M 7LH

Patents ADP number *(if you know it)*

745002

6. If you are declaring priority from one or more earlier patent applications, give the country and the date of filing of the or of each of these earlier applications and *(if you know it)* the or each application number

Country	Priority application number *(if you know it)*	Date of filing *(day / month / year)*

7. If this application is divided or otherwise derived from an earlier UK application, give the number and the filing date of the earlier application

Number of earlier application	Date of filing *(day / month / year)*

8. Is a statement of inventorship and of right to grant of a patent required in support of this request? *(Answer 'Yes' if:*

a) any applicant named in part 3 is not an inventor, or
b) there is an inventor who is not named as an applicant, or
c) any named applicant is a corporate body.
See note (d))

YES

Patents Form 1/77

9. Enter the number of sheets for any of the following items you are filing with this form. Do not count copies of the same document

Continuation sheets of this form	
Description	13
Claim(s)	2
Abstract	
Drawing(s)	14

10. If you are also filing any of the following, state how many against each item.

Priority documents	
Translations of priority documents	
Statement of inventorship and right to grant of a patent *(Patents Form 7/77)*	
Request for preliminary examination and search *(Patents Form 9/77)*	
Request for substantive examination *(Patents Form 10/77)*	
Any other documents *(please specify)*	NO

11. For the applicant
Gill Jennings & Every

I/We request the grant of a patent on the basis of this application.

Signature [signature] Date 16 April 2003

12. Name and daytime telephone number of person to contact in the United Kingdom

JAPPY, John William Graham
020 7377 1377

Warning

After an application for a patent has been filed, the Comptroller of the Patent Office will consider whether publication or communication of the invention should be prohibited or restricted under Section 22 of the Patents Act 1977. You will be informed if it is necessary to prohibit or restrict your invention in this way. Furthermore, if you live in the United Kingdom, Section 23 of the Patents Act 1977 stops you from applying for a patent abroad without first getting written permission from the Patent Office unless an application has been filed at least 6 weeks beforehand in the United Kingdom for a patent for the same invention and either no direction prohibiting publication or communication has been given, or any such direction has been revoked.

Notes

a) *If you need help to fill in this form or you have any questions, please contact the Patent Office on 0645 500505.*

b) *Write your answers in capital letters using black ink or you may type them.*

c) *If there is not enough space for all the relevant details on any part of this form, please continue on a separate sheet of paper and write "see continuation sheet" in the relevant part(s). Any continuation sheet should be attached to this form.*

d) *If you have answered 'Yes' Patents Form 7/77 will need to be filed.*

e) *Once you have filled in the form you must remember to sign and date it.*

f) *For details of the fee and ways to pay please contact the Patent Office.*

METHOD

Field of the Invention

This invention relates to methods for converting a target polynucleotide into readable information chains, for subsequent determination of the sequence of the target polynucleotide.

Background to the Invention

Advances in the study of molecules have been led, in part, by improvement in technologies used to characterise the molecules or their biological reactions. In particular, the study of the nucleic acids DNA and RNA has benefited from developing technologies used for sequence analysis and the study of hybridisation events.

The principal method in general use for large-scale DNA sequencing is the chain termination method. This method was first developed by Sanger and Coulson (Sanger *et al.*, Proc. Natl. Acad. Sci. USA, 1977; 74: 5463-5467), and relies on the use of dideoxy derivatives of the four nucleotides which are incorporated into the nascent polynucleotide chain in a polymerase reaction. Upon incorporation, the dideoxy derivatives terminate the polymerase reaction and the products are then separated by gel electrophoresis and analysed to reveal the position at which the particular dideoxy derivative was incorporated into the chain.

Although this method is widely used and produces reliable results, it is recognised that it is slow, labour-intensive and expensive.

US-A-5302509 discloses a method to sequence a polynucleotide immobilised on a solid support. The method relies on the incorporation of 3-blocked bases A, G, C and T having a different fluorescent label to the immobilised polynucleotide, in the presence of DNA polymerase. The polymerase incorporates a base complementary to the target polynucleotide, but is prevented from further addition by the 3'-blocking group. The label of the incorporated base can then be determined and the blocking group removed by chemical cleavage to allow further polymerisation to occur. However, the need to remove the blocking groups in this manner is time-consuming and must be performed with high efficiency.

WO-A-00/39333 describes a method for sequencing polynucleotide by converting the sequence of a target polynucleotide into a second polynucleotide having a defined sequence and positional information contained therein. The sequence information of the target is said to be "magnified" in the second polynucleotide, allowing greater ease of distinguishing between the individual bases on the target molecule. This is achieved using "magnifying tags" which are predetermined nucleic acid sequences. Each of the bases adenine, cytosine, guanine and thymine on the target molecule is represented by an individual magnifying tag, converting the original target sequence into a magnified sequence. Conventional techniques may then be used to determine the order of the magnifying tags, and thereby determining the specific sequence on the target polynucleotide.

In a preferred sequencing method, each magnifying tag comprises a label, e.g. a fluorescent label, which may then be identified and used to characterise the magnifying tag.

Although the method disclosed in this patent publication has many advantages, there is still a need for improved methods for sequencing target polynucleotides.

Summary of the Invention

According to a first aspect of the invention, a method for determining the sequence of a target polynucleotide, comprises the steps of:

i) treating a sample of a double-stranded target polynucleotide to create overhangs at the 5' and 3' ends, each with a defined number of bases;

ii) dividing the sample and contacting each separate sample with a double stranded polynucleotide signal sequence and a double stranded adapter polynucleotide, each signal sequence representing a specific polynucleotide sequence of the same length as that of the 5'-overhang, and comprising an overhang that permits hybridisation and ligation to the 3' end of the target polynucleotide, and each adapter comprising an overhang that is

of complementary sequence to that of the sequence represented by the signal sequence;

iii) carrying out the polymerase reaction on the sample(s) using primers that hybridise at the end of the signal sequence and adapter sequence, wherein the product of the polymerase reaction comprises a restriction site that permits cleavage of the adapter and that part of the target polynucleotide that formed the 5'-overhang, to form a new 5'-overhang;

iv) repeating steps (i) to (iii) using restriction enzymes to create the overhangs; and

v) identifying which signal sequences are present on the amplified products, and in which order, to thereby determine the sequence of the target polynucleotide.

Description of the Invention

The method of the present invention is an adaptation of the conversion method disclosed in WO-A-00/39333. In summary, a defined number of sequences at one end of the target polynucleotide is converted into a defined signal sequence at the other end of the target polynucleotide. Through a series of steps, the "sequenced" end of the target is cleaved, while the signal sequence end is built up, generating a defined series of signal sequences which can be determined through a subsequent identification step.

The term "polynucleotide" is well known in the art and is used to refer to a series of linked nucleic acid molecules, e.g. DNA or RNA. Nucleic acid mimics, e.g. PNA, LNA (locked nucleic acid) and 2'-O-methRNA are also within the scope of the invention.

The reference to the bases A, T(U), G and C, relate to the nucleotide bases adenine, thymine (uracil), guanine and cytosine, as will be appreciated in the art. Uracil replaces thymine when the polynucleotide is RNA, or it can be introduced into DNA using dUTP, again as well understood in the art.

Reference is made herein to the 5' and 3' ends of the target polynucleotide. As will be apparent, the target polynucleotide is double stranded and so each end comprises both 5' and 3' ends. However, the reference to 5'

and 3' ends is made with specific reference to the strand of the target that is being sequenced. It will also be apparent that while the specification refers to the 5' end as the end being sequenced, it will be possible to modify the procedure so that it is the 3' end that is sequenced, i.e. the signal sequences and adapters ligate to opposite ends to those indicated herein.

The method of the invention is carried out to identify a defined number of bases, e.g. four bases, at one end of a target polynucleotide, and then convert this information into a signal sequence on the opposite end, thereby permitting additional conversion cycles to take place, without the removal of incorporated signal sequences from previous cycles.

The method occurs by obtaining the target polynucleotide and treating this with a series of restriction enzymes to form defined overhang regions at each end of the polynucleotide, which allow incorporation of a signal sequence and an adapter during each conversion step.

A "signal sequence" is a double-stranded polynucleotide that comprises distinct "units" of nucleic acid sequence. Each of the bases A, T(U), G and C on the target is represented by a distinct and predefined unit, or unique combination of units. Each unit will preferably comprise two or more nucleotide bases, preferably from 2 to 50 bases, more preferably 2 to 20 bases and most preferably 4 to 10 bases, e.g. 6 bases. There are at least two different bases contained in each unit. In a preferred embodiment, there are three different bases in each unit. The design of the units is such that it will be possible to distinguish the different units during a "read-out" step, e.g. involving either the incorporation of detectably labelled nucleotides in a polymerisation reaction, or on hybridisation of complementary oligonucleotides. For example, each base on the target is represented by a series of bases in a unit, where one base will be complementary to a labelled nucleotide introduced during the read-out step, one base will act as a "spacer" to provide separation between incorporated labels, and one base will act as a stop signal.

In a preferred embodiment, each signal sequence comprises two units of distinct sequence which represent all of the four bases on the target. According to this embodiment, the two units can be used as a binary system, with one unit

representing "0" and the other representing "1". Each base on the target is characterised by a combination of the two units. For example, adenine may be represented by "0" + "0", cytosine by "0" + "1", guanine by "1" + "0" and thymine by "1" + "1", as shown in Figure 1. It is necessary to distinguish between the units, and so a "stop" signal can be incorporated into each unit. It is also preferable to use different units representing "1" and "0", depending on whether the base on the target (template) polynucleotide is in an odd or even numbered position.

This is demonstrated as follows:

Odd numbered template sequence:

"0" : ATTTTTAT(CC)

"1" : GTTTTTGT(CC)

Even numbered template sequence:

"0" : ACCCCCAC(TT)

"1" : GCCCCCGC(TT)

Suitable signal sequences are also described in WO-A-00/39333.

The "adapter" is a double stranded polynucleotide that comprises a defined overhang that is capable of hybridising and ligating to a complementary overhang on the target polynucleotide. The adapter is designed to hybridise to that part of the target polynucleotide that is being sequenced, i.e. the 5'-overhang part of the target polynucleotide. If the 5'-overhang on the target is 4 bases, then the complementary overhang on the adapter will also be 4 bases. This permits ligation to occur to form a double-stranded polynucleotide. Given that the 5'-overhang of the target is of unknown sequence, it will usually be necessary to use a combination of adaptors having all permutations of the (for example) 4 base sequence. The adapters of different sequence will be used in separate reactions, as discussed below. The adapters will usually have restriction enzyme recognition sites incorporated, allowing their cleavage during a later step, to create a new 5'-overhang.

The method comprises the following general steps:

Digestion

A sample of the target polynucleotide is first treated with one or more restriction enzymes, resulting in an overhang in the target with a defined number of bases, e.g. 3 bases, that is used for ligation of signal sequences later in the procedure (the overhang). An overhang is also created at the opposite end of the target polynucleotide (the 5'-overhang). It is this overhang that is to be identified (represented) by the signal sequence during the conversion process and is to be ligated to the adapter.

Ligation

The sample with the target polynucleotide is divided evenly into reaction compartments representing all permutations of the overhang to be sequenced. For example, if the overhang is 4 bases, then 256 reaction compartments are to be used. In the first compartment, the target polynucleotide is contacted with a signal sequence that represents for example AAAA (when the overhang to be sequenced is 4 bases), the signal sequence being ligated to the 3'-overhang on all the target polynucleotides. An adapter polynucleotide with the overhang TTTT is also added, but this will only ligate specifically to the target polynucleotides which contain an AAAA overhang, leaving polynucleotides with a different overhang sequence with an unligated overhang. The same procedure is performed in the other compartments, each representing a different combination of the possible 4 base sequence. In each case the adapter will be the complement of the sequence represented by the signal sequence.

Amplification

After the ligation steps have been carried out, the samples can be pooled, and a polymerase reaction carried out. The polymerase reaction is carried out using primers which target the ends of the signal sequence and adapter, and therefore only those molecules that have successfully ligated both the signal sequence and the adapter sequence will be amplified exponentially, while those molecules containing only a signal sequence will be removed as a consequence of linear amplification. The result will be a population of converted polynucleotide fragments, where the 4 base overhang that was initially generated at one end of the target polynucleotide has been replaced with a

signal sequence representing the 4 bases at the other end of the target polynucleotide.

In a preferred embodiment, the polymerase reaction is carried out using methyl-dCTP, which ensures that native restriction enzyme sites remain inactive. The polymerase reaction is also carried out using specific primers which incorporate restriction enzyme sites within their sequence, thereby allowing further conversion cycles to take place. The primers are specific for sequences found within the signal sequence and adapter, thereby ensuring that the amplification step occurs only when both the signal sequence and adapter is ligated onto the target polynucleotide.

The overhangs may be produced by different means, although the preferred embodiment is to use restriction enzymes, e.g. class IIs restriction enzymes. These enzymes exhibit no specificity to the sequence that is cut and they can therefore generate overhangs with all types of base compositions. The binding site of the restriction enzyme can be located so that an overhang is formed inside the actual target polynucleotide. In practice, it is preferable to choose enzymes that generate 3-4 base pair overhangs.

Class IIs restriction endonucleases are known, and are identified in WO-A-00/39333.

In a preferred embodiment, the target polynucleotide is first treated with an adapter polynucleotide which incorporates a recognition sequence for the restriction enzyme BbvI, and the polynucleotide is then treated with the restriction enzyme to generate a specific overhang at the 5' end. Native recognition sites for BbvI are inactivated by first treating the target polynucleotide by methylation. Having created the first overhang, subsequent overhangs at the 5' end are preferably created using the restriction enzyme *Sfa*NI. This is carried out by ligating a specific adapter sequence at the 5' end which incorporates the recognition sequence for *Sfa*NI.

The initial overhang at the 3' end is preferably created by ligating a polynucleotide that incorporates a defined overhang. The defined overhang permits a first signal sequence to hybridise and ligate at the 3' end. The signal sequence incorporates a restriction enzyme site, which is preferably for the

restriction enzyme *EarI*. Subsequent cycles of cleavage and incorporation occur using the restriction enzymes *EarI* and *Sfa*NI, as shown in Figure 3.

The conversion cycle is further illustrated in Figure 4, which shows the fate of five different fragments within three arbitrarily picked reaction compartments out of the 256 compartments used in the procedure. The fragments are first evenly distributed into the 256 compartments. In compartment 56, where the overhang "TCTA" is to be identified, a signal sequence representing this base composition is introduced and ligated with the 3 base 3'-overhang on the left end of the target polynucleotide. This ligation occurs regardless of the composition of the 4 base 5'-overhang on the right side of the polynucleotide. A specific adapter that only ligates with fragments with 3' TCTA -5' overhangs is then introduced and ligated. A final amplification reaction then selectively amplifies the fragments that have ligated with the adapter and hence removes the other fragments which have been associated with an incorrect signal sequence. The same procedure is performed in compartments 141 and 194 with the exception that the composition of the signal sequences and the overhang adapters are adjusted in accordance with the overhangs that are to be identified in the wells (3'-ATAG-5' and 3'-CAAT-5').

The target polynucleotides from the 256 compartments are preferably pulled into one common reaction tube before the amplification reaction takes place, thereby avoiding the inconvenience of performing at 256 separate amplification reactions.

After subsequent conversion cycles are carried out, the sequence at the 5' end of the target polynucleotide is reduced and the signal sequences introduced at the 3' end of the target polynucleotide are increased in a defined way corresponding to the now "sequenced" target polynucleotide. Determining the type and order of the signal sequences may be carried out using methods disclosed in WO-A-00/39333, or in the co-pending British Patent Application filed on 16 April 2003 in the name of LingVitae.

The following Example illustrates the invention.

Example

The principle of cyclic conversion of nucleic acid bases into readable signal chains was demonstrated by converting a sequence of 12 bases in a target DNA into its corresponding signal chain. The conversion took place in steps through a cyclic method, where 4 bases were converted in each cycle for a total number of 3 cycles. The end product was a signal chain consisting of 12 signal components representing the relevant base sequence in the target DNA. The experiment is described schematically in Figure 3.

A target DNA fragment of 240 bp, containing internal sites for *Sfa*NI and *Eam*1104I was PCR-amplified from the bacteriophage Lambda genome (1507-1703) using primers containing restriction sites for *Eam*1104I and *Sfa*NI (the top fragment in Figure 3) respectively. In order to avoid cutting from the internal sites later, the PCR reaction was conducted with methylated deoxy-Cytosine nucleotides (m5-dCTP) instead of the usual deoxy-Cytosin (dCTP). PCR conditions: (50 µl): 10 mM KCl, 10mM $(NH_4)_2SO_4$, 20mM Tris-HCl, 0.1% Triton-X-100, pH 8.8 (Thermopol buffer, New England Biolabs), 2 mM Mg^{2+}, 200 µM dNTP (- dCTP) and 200 µM m5-dCTP (Amersham Pharmacia Biotech), 20 pmol Lambda primer #328 (5'-agactggcgatccctggcatcacccctccagcgtgttttat-3') and 20 mpol Lambda primer #329 (5'-gcactgataggcgtcactcttcgctgtacgctgtccagatgt-3') (MWG biotech), 10 ng Lambda genome, 1U Vent polymerase (New England Biolabs). The PCR cycling was conducted with a PTC-200 (MJResearch). Hot start: 95°C, 5 minutes, 35 cycles consisting of: 95°C, 15 seconds, 58°C, 20 seconds, 72°C, 30 seconds. Complete extension step: 72°C, 5 minutes.

Cycle 1:

The fragment enters the sequencing cycle (Figure 3, A) and is cut with *Eam*1104I and *Sfa*NI (as the internal sites are methylated and protected against cutting, cutting will only take place from the primer areas). Digestion conditions: 100 mM NaCl, 50 mM Tris-HCl, 10 mM $MgCl_2$, 1 mM DTT, pH 7.9 (NEB3, New England Biolabs), 1 µg methylated PCR fragment, 20 U *Eam*1104I, 3U *Sfa*NI, and incubation for one hour at 37°C. The enzymes were heat-inactivated at 65°C for 20 minutes. The cut fragment was cleaned using GibcoBRL PCR purification system (Gibco) and then elated with 10 mM Tris-HCl. The cut fragment corresponds to the nature of those fragments that are obtained after an

initial round of the sequencing method (described in the main patent) in that it has an overhang of 3 bases for signal chain ligation and an overhang of 4 bases for ligation to a specific right hand adapter.

The conversion step where the selection marker is attached (Figure 3, B) was carried out by ligating the fragment with the specific right hand adapter and its associated signal chain: 40 mM Tris-HCl, 10 mM $MgCl_2$, 10 mM DTT, 0.5 mM ATP, pH 7.8 (Ligase buffer, Fermentas), 1.6 pmol right hand adapter1 (5'-Biotin-ggctaggtgctgatgaacgcatcg-3' annealed to 5'-tggacgatgcgttcatcagcacctagcc-3') (MWG-biotech), 1.9 pmol signal chain1 (1), 7.5 Weiss-U T4 DNA ligase (Fermentas). The incubation took place at room temperature (22°C) for 2.5 hours. The ligase was heat-inactivated at 65°C for 10 minutes.

The signal chains used in the experiment were made in the following way: Signal components (each representing one base) were made by annealing oligonucleotides (approx. 45 bases) which again were ligated together to create the desired signal chains (the signal components are constructed with different overhangs which dictate the sequence of ligation). The signal chains were amplified by PCR and cut with *Sap*I in order to generate the complementary overhang necessary for ligation of the signal chain to the target DNA. Since all target DNA fragments have the same overhang for signal chain ligation, all fragments in a given reaction will be ligated to the same signal chain, irrespective of the sequence of the *Sfa*NI-generated overhang. Specificity and selection are therefore located in the right hand adapter. As this ligates to that terminus of the target DNA which was cut with *Sfa*NI, only fragments with complementary overhang to the right hand adapter will be ligated. A conversion therefore presupposes that the target DNA, as well as being ligated to a chain, also has an overhang complementary to the right hand adapter in order for ligation to this to take place. In other words, the right hand adapter ensures a specific selection of fragments with an overhang sequence corresponding to the ligated signal chain.

The selection step of the base conversion consists of selection and amplification of fragments which have had ligated to them both a signal chain and a specific right hand adapter (Figure 3,C). This was conducted using PCR

under the following conditions (50 µl): 10 mM KCl, 10 mM $(NH_4)_2SO_4$, 20 mM Tris-HCl, 0.1% Triton-X-100, pH 8.8 (Thermopol buffer, New England Biolabs), 2 mM Mg^{2+}, 200 µM dNTP (-dCTP) and 200 µM m5-dCTP (Amersham Pharmacia Biotech), 20 pmol right hand adapter1-primer #332 (5'-Biotin-ggctaggtgctgatgaacgcatcg-3') and 20 pmol Signal chain1-primer #340 (5'-taatacgactcactatagcatgactcgagcctcttcgcga-3') (MWG-biotech), approx. 3.5 fmol ligated target DNA and 1U Vent polymerase (New England Biolabs). PCR cycling was conducted with a PTC-200 (MJResearch). Hot start: 95°C, 2 minutes, 20 cycles consisting of: 95°C, 15 seconds, 66°C, 20 seconds, 72°C, 30 seconds. Complete extension step: 72°C, 5 minutes. Well 2 in Figure 2 shows the results of the initial cycle of the sequencing method. The correct fragment of 380 bp was generated.

Cycle 2:

In order to convert the next 4 bases, the PCR product was cleaned with the GibcoBRL purification system (Gibco) and then cut with *Sfa*NI and *Eam*1104I. It is possible for this to take place because the signal chain and the right hand adapter from the initial cycle contain a site for *Eam*1104I and *Sfa*NI respectively. As the sites were located in the primer region, they are not blocked by methylation during PCR. To increase the efficiency of the cutting reaction, the cuts were conducted serially under optimal cutting conditions. *Sfa*NI cutting: 100 mM NaCl, 50 mM Tris-HCl, 10 mM $MgCl_2$, 1 mM DTT, pH 7.9 (NEB3, New England Biolabs), 1 µg methylated PCR fragment, 4U *Sfa*NI. Incubation at 37°C for 1 hour. Heat inactivation of the enzyme at 65°C for 10 minutes. A Micro Bio-Spin 6 column (BioRad) was used to clean the fragment. *Eam*1104I cutting: 33 mM Tris-acetate, 10 mM magnesium-acetate, 66 mM potassium-acetate, 0.1 mg/ml BSA, pH 7.9 (Tango Y^+, Fermentas), *Sfa*NI-digested PCR fragment, 10U *Eam*1104I. Incubated at 37°C for 1 hour, followed by heat inactivating at 65°C for 20 minutes. The digested product was cleaned by using the GibcoBRL purification system (Gibco). The initial step of the conversion was conducted corresponding to the initial cycle, except that a new signal chain and its associated specific right hand adapter (corresponding to the next 4 bases of the sequence) were added. To reduce potential carry-over problems, signal chain2

was designed with an different overhang from signal chain1. Ligation conditions: 40 mM Tris-HCl, 10 mM $MgCl_2$, 10 mM DTT, 0.5 mM ATP, pH 7.8 (Ligase buffer, Fermentas), 0.5 pmol right hand adapter2 (5'Biotin-cgacagtacgacggaccagcatcc-3' annealed to 5'-acgcggatgctggtccgtcgtactgtcg-3') (MWG-biotech), 1 pmol signal chain2 (2), 1 Weiss-U T4 DNA ligase (Fermentas). The incubation took place at room temperature (22°C) for 1.5 hours. The ligase was heat-inactivated at 65°C for 10 minutes. PCR amplification took place under the same conditions as the initial cycle, except that a new primer set (right hand adapter2-primer #347, 5'-Biotin-cgacagtacgacggaccagcatcc-3' and signal chain2-primer #343, 5'-taatacgactcactatagcatcgaatgaccgcctcttccact-3'). Different primer sets for each sequencing cycle are preferable in order to minimise the danger of amplifying any remains from the previous sequencing cycle. The result of the PCR amplification in the second cycle is shown in Well 3 in Figure 2. The correct fragment of 523 bp was generated.

Cycle 3.

Conversion of the next 4 bases and the PCR amplification followed the same pattern as described for Cycle 2: After a serial cutting with *Sfa*NI and *Eam*1104I, the fragment was ligated with signal chain3 (3) and right hand adapter3 (5'-Biotin-atcgagcctggcatagcagcatca-3' annealed to 5'-aaactgatgctatgccaggctcgat-3) (MWG-Biotech). The PCR amplification was conducted with the primer set: the right hand adapter3-primer #353 (5'-atcgagcctggcatagcagcatca-3') and signal chain3-primer #345 (5'-taatacgactcactatagcaccgggcaggatagactcttcaggt-3'). The result of the amplification in Cycle 3 is shown in Well 4 in Figure 5. Well 4 shows that two weak and one relatively strong bands are formed. The strong band comes closest to the expected length of 666 bp. The weak bands of the wrong size which can be seen in Well 4 (Cycle 3) and Well 3 (Cycle 2) are most probably the result of carry-over problems and mispriming. The weak band in Well 3 and the weak bands further down in Well 4 correspond to the size of the PCR fragment in the previous cycle. A possible explanation is that incomplete cutting of fragments in one cycle may function as a template in the next PCR cycle. Even though new primer sets in each cycle reduce this problem somewhat, the

danger of mispriming is still present because of the enzyme sites the primers have in common. However, this type of mispriming may be eliminated by using more stringent annealing conditions during PCR (e.g. Mg^{2+}, temperature), the use of more discriminating polymerases, moving the *Eam*1104I site away from 3' on the sequence chain primer, choice of new sequences or immobilisation of the fragments (e.g. the biotin-streptavidin system on microbeads) before cutting so that those fragments entering the next cycle are guaranteed to have been cut. The best solution for other types of mispriming would be to choose new primer sequences and/or optimise PCR conditions. The correct fragments in Cycles 1-3 in the example demonstrate the feasibility of a cyclic conversion of DNA bases into readable signal chains.

CLAIMS

1. A method for determining the sequence of a target polynucleotide, comprising the steps of:

i) treating a sample of a double-stranded target polynucleotide to create overhangs at each end, one of which is to be sequenced, each overhang having a defined number of bases;

ii) dividing the sample and contacting each separate sample with a double stranded polynucleotide signal sequence and a double stranded adapter polynucleotide, each signal sequence representing a specific polynucleotide sequence of the same length as that of the overhang to be sequenced and comprising an overhang that permits hybridisation and ligation to the end of the target polynucleotide opposite that being sequenced, and each adapter comprising an overhang that is of complementary sequence to the overhang sequence being sequenced;

iii) carrying out the polymerase reaction on the sample(s) using primers that hybridise at the ends of the signal sequence and adapter sequence, wherein the product of the polymerase reaction comprises a restriction site that permits cleavage of the adapter to form a new overhang to be sequenced;

iv) repeating steps (i) to (iii) using restriction enzymes to create the overhangs; and

v) identifying which signal sequences are present on the amplified products, and in which order, to thereby determine the sequence of the target polynucleotide.

2. A method according to claim 1, wherein the restriction enzymes are class IIs enzymes.

3. A method according to claim 1 or claim 2, wherein the restriction enzymes are *Sfa*NI and *Ear*I.

4. A method according to any preceding claim, wherein the polymerase reaction is carried out using methyl-dCTP as a replacement for dCTP.

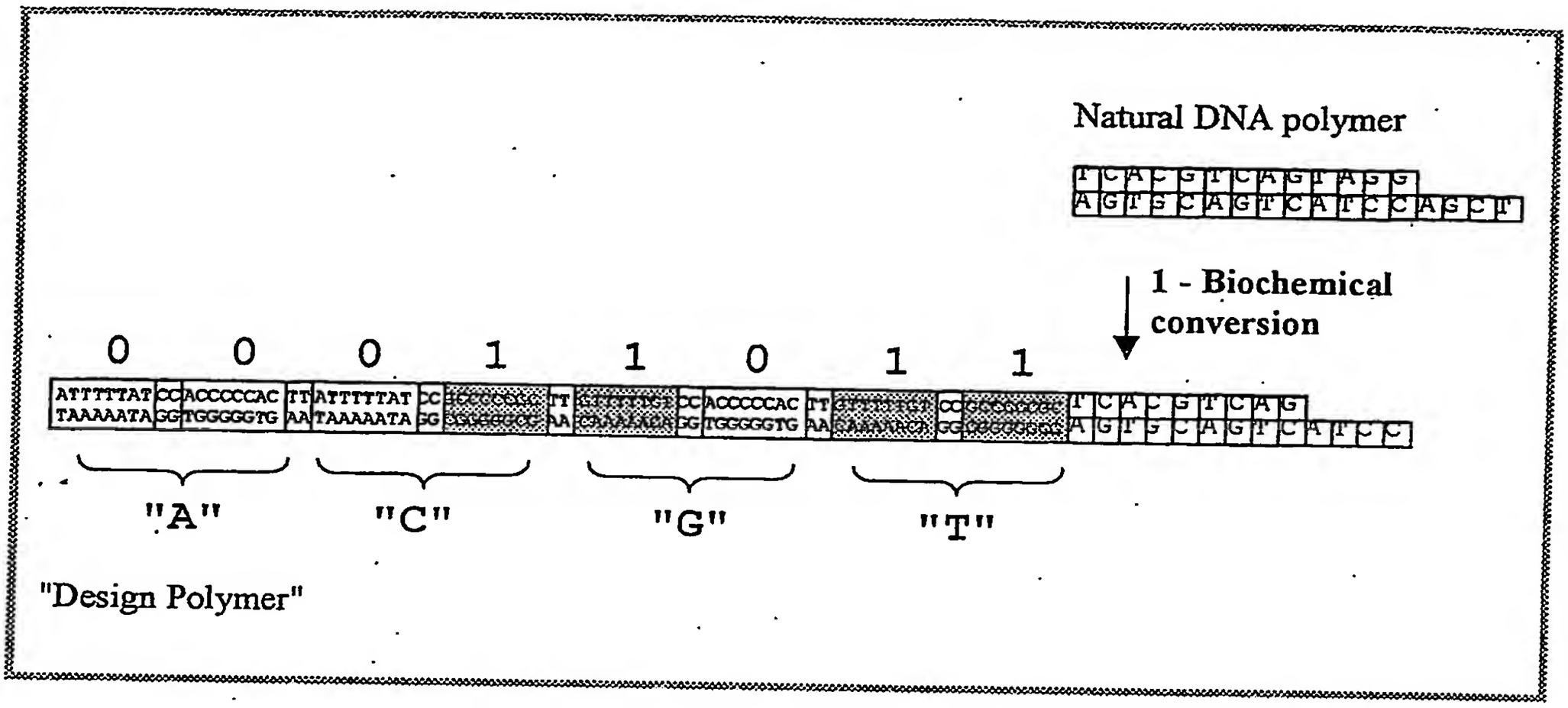
5. A method according to any preceding claim, wherein the overhang that

ligates to the signal sequence is 3 bases.

6. A method according to any preceding claim, wherein the overhang to be sequenced comprises 4 bases.

7. A method according to any preceding claim, wherein the adapter is immobilised on a support material.

Figure 1



"Design Polymer"

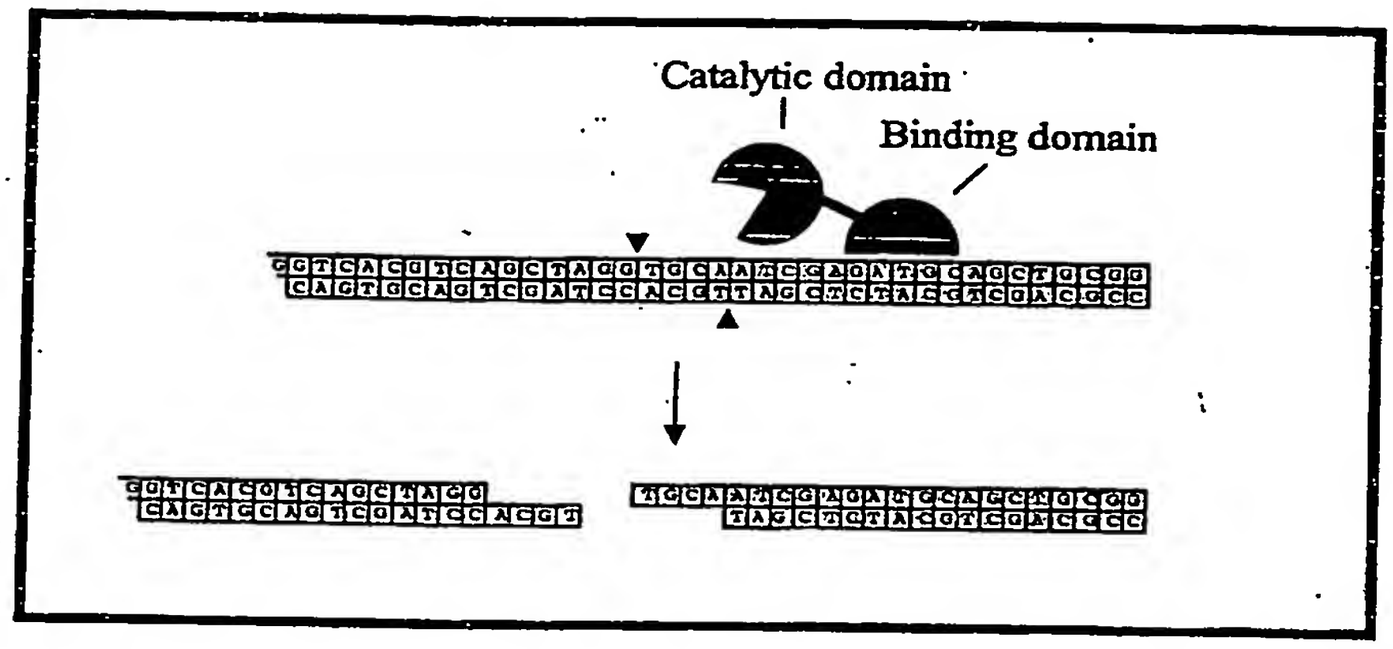
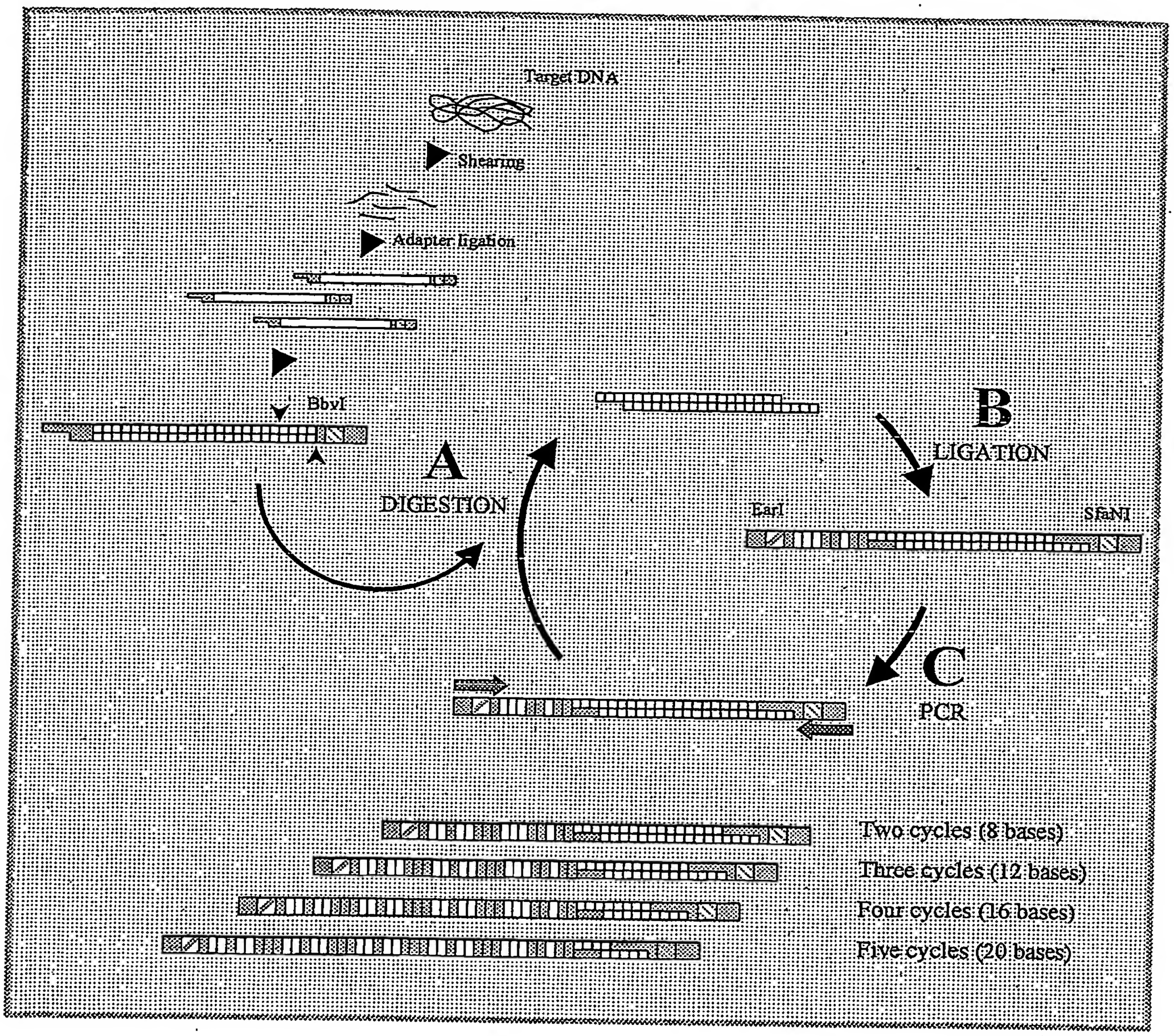


Figure 2

Figure 3


Target DNA
Shearing
Adapter ligation
BbvI
N
A
DIGESTION
B
LIGATION
EarI
SfaNI
C
PCR
Two cycles (8 bases)
Three cycles (12 bases)
Four cycles (16 bases)
Five cycles (20 bases)

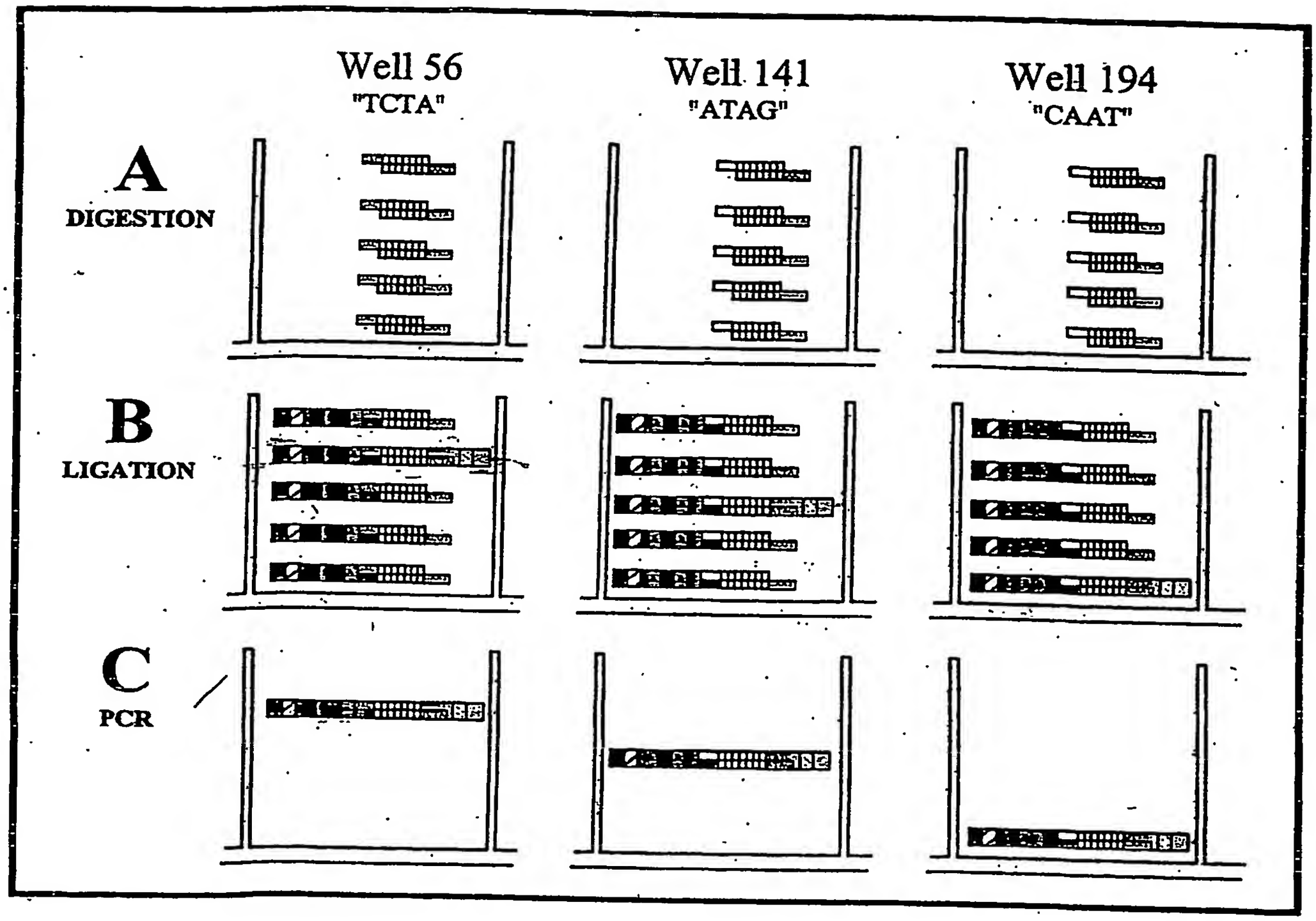
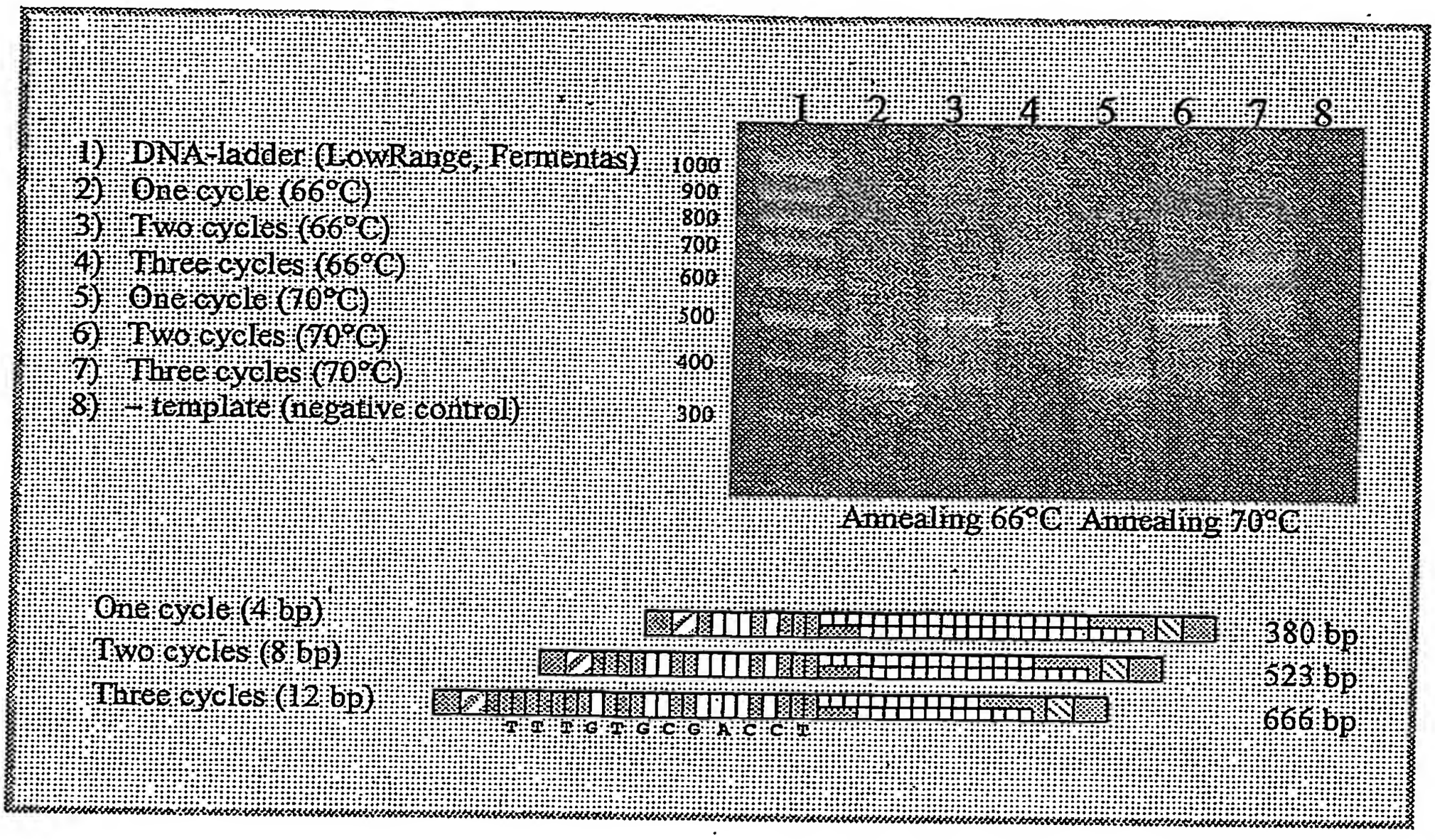

Well 56
"TCTA"
Well 141
"ATAG"
Well 194
"CAAT"
A
DIGESTION
B
LIGATION
C
PCR

Figure 4

Figure 5


1) DNA-ladder (LowRange, Fermentas)
2) One cycle (66°C)
3) Two cycles (66°C)
4) Three cycles (66°C)
5) One cycle (70°C)
6) Two cycles (70°C)
7) Three cycles (70°C)
8) – template (negative control)
1 2 3 4 5 6 7 8
1000
900
800
700
600
500
400
300
Annealing 66°C Annealing 70°C
One cycle (4 bp)
Two cycles (8 bp)
Three cycles (12 bp)
T T T G T G C G A C C T
380 bp
523 bp
666 bp

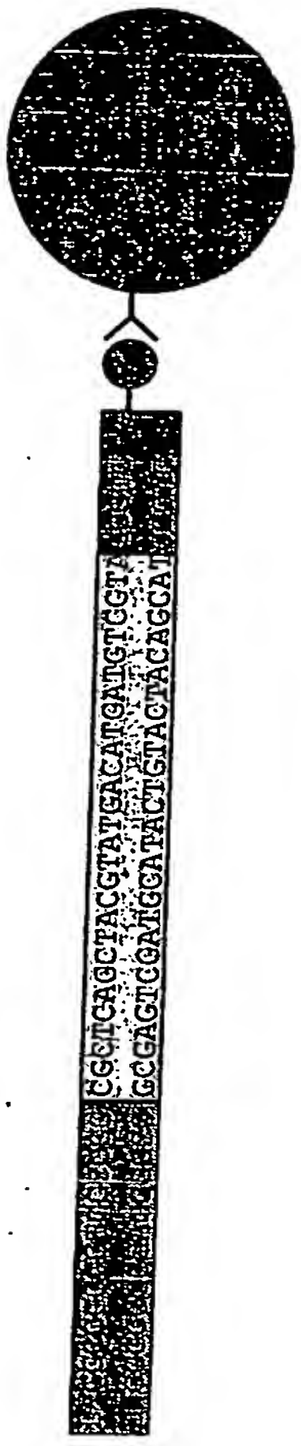
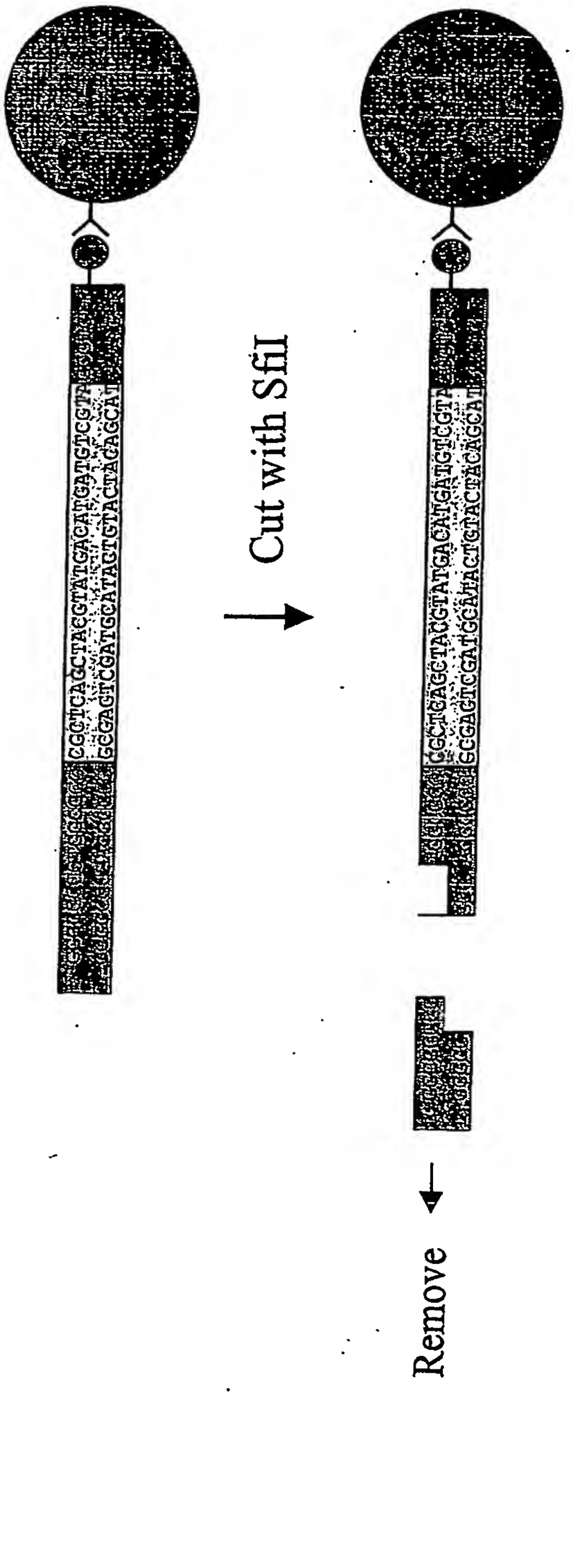
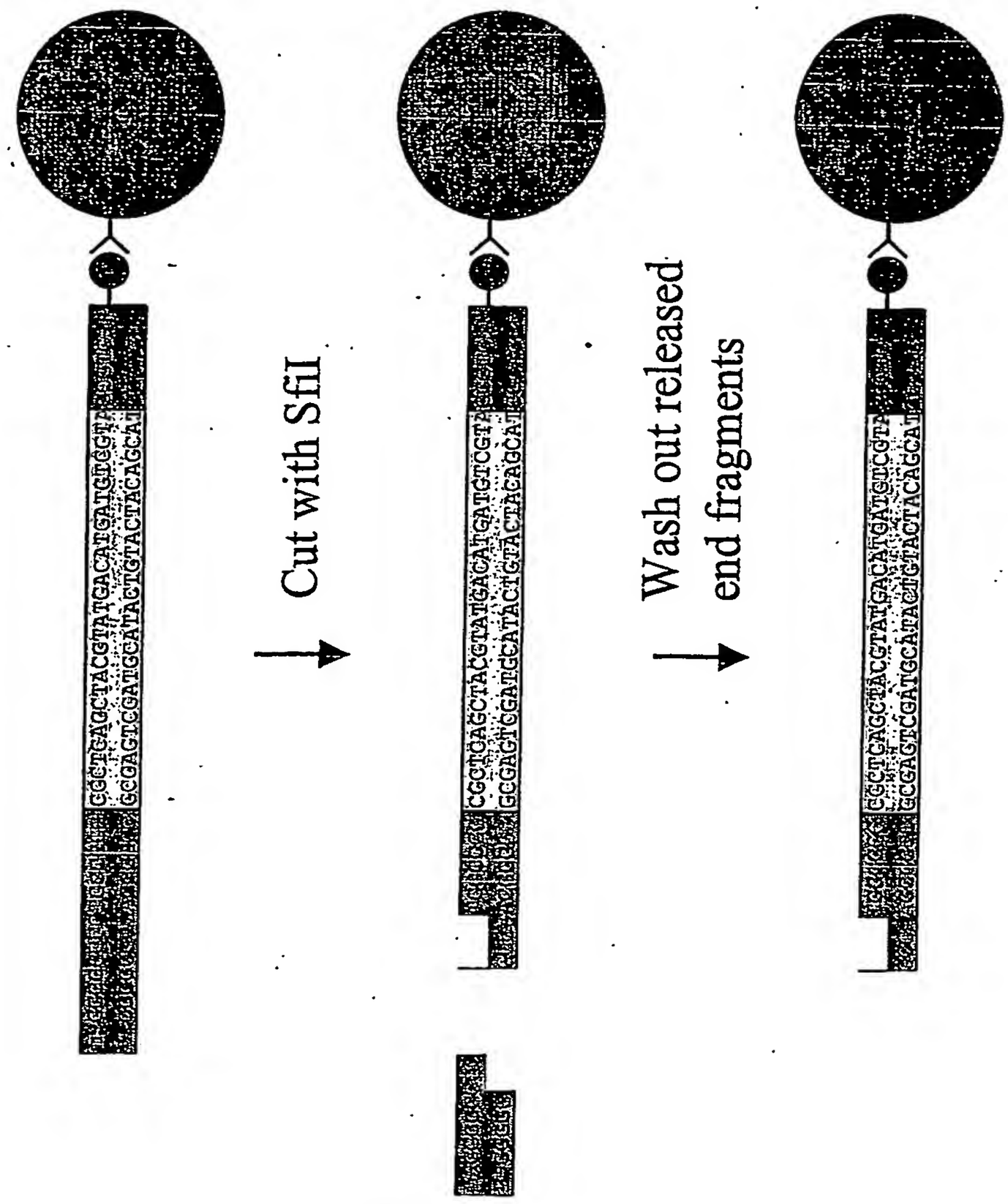
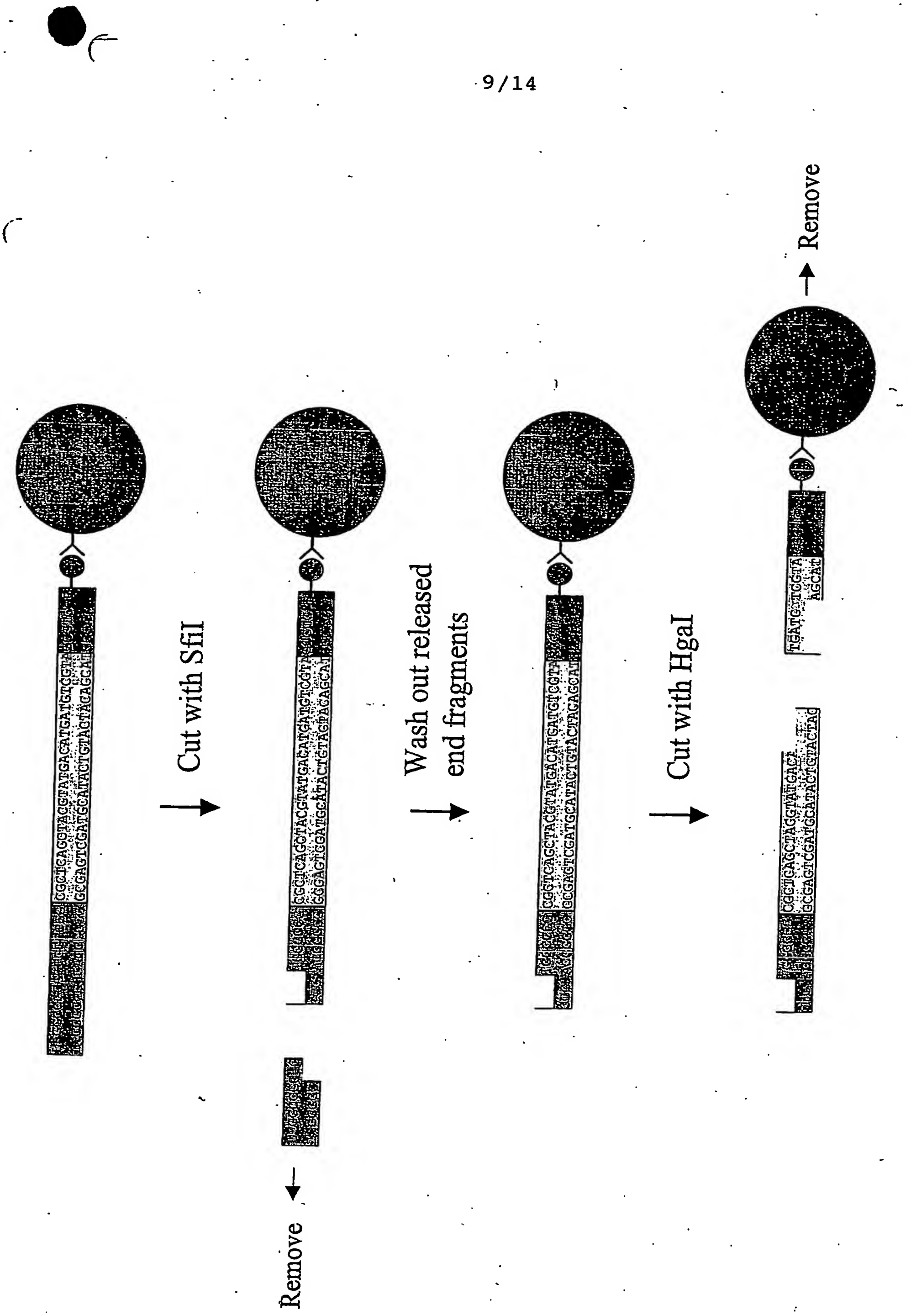
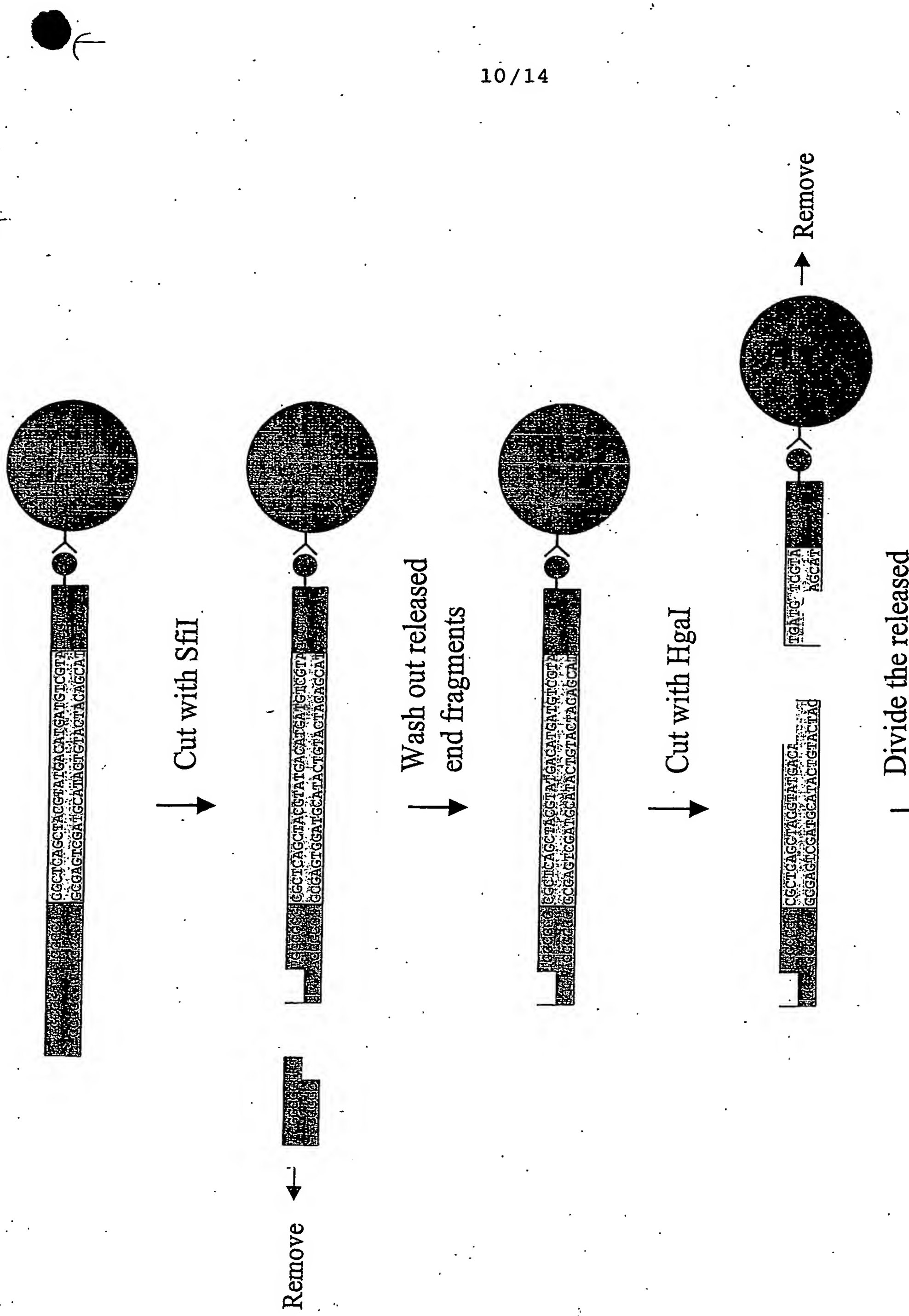

GGCTCAGCCTACGTATGACATGATGGTA
GCGAGTCGATGCATACTGTACTACAGCAT

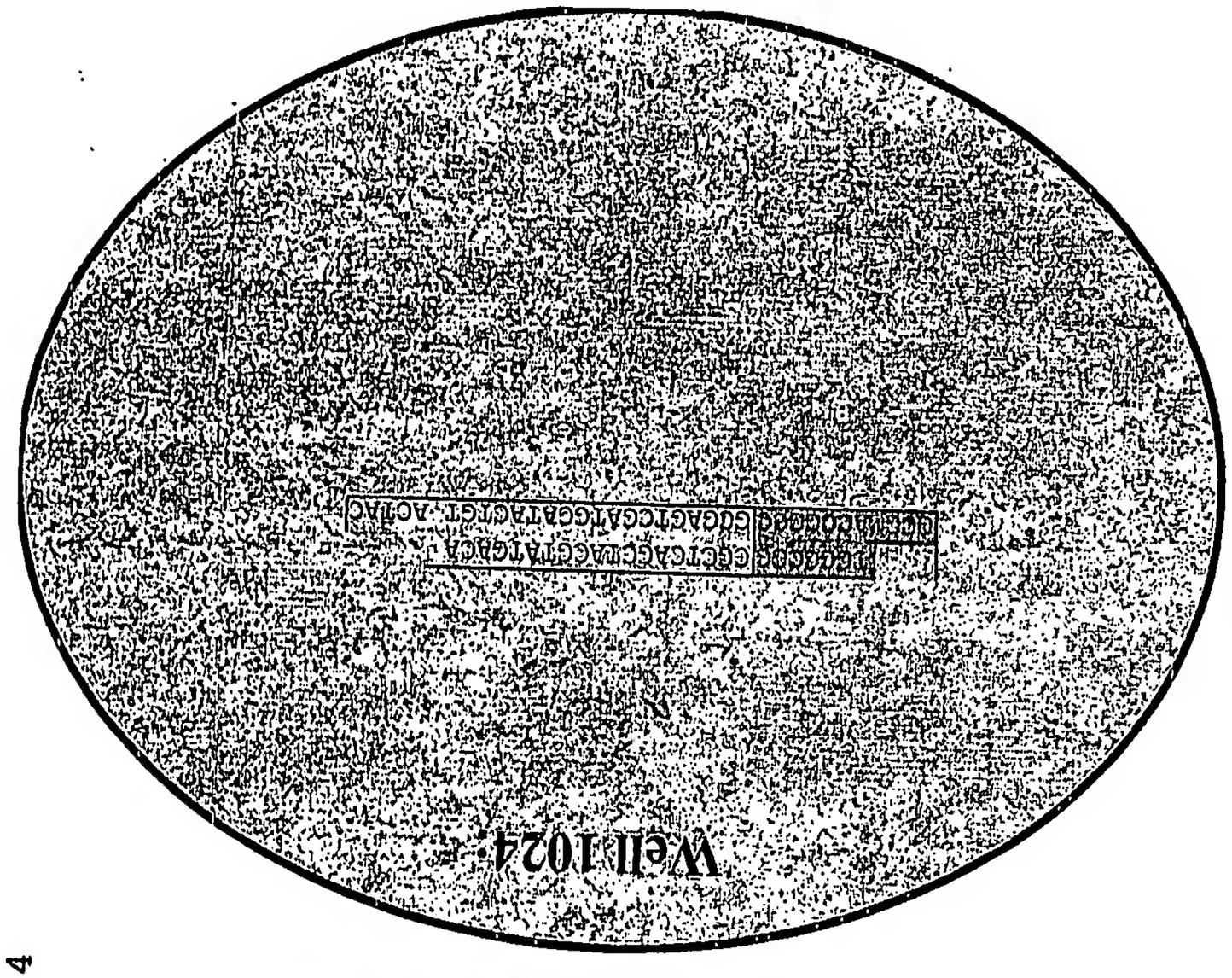
Figure 6

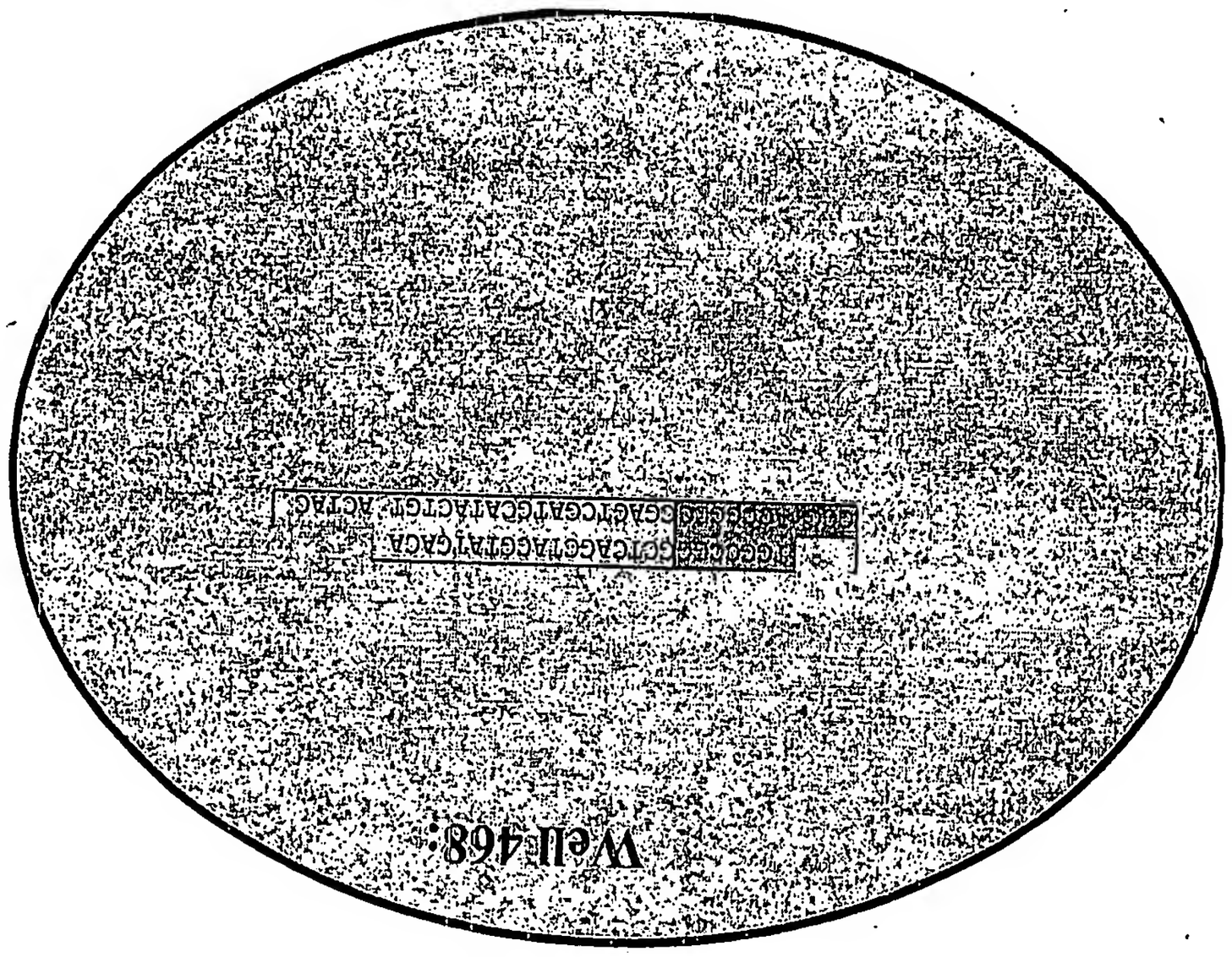

Cut with SfiI
Remove

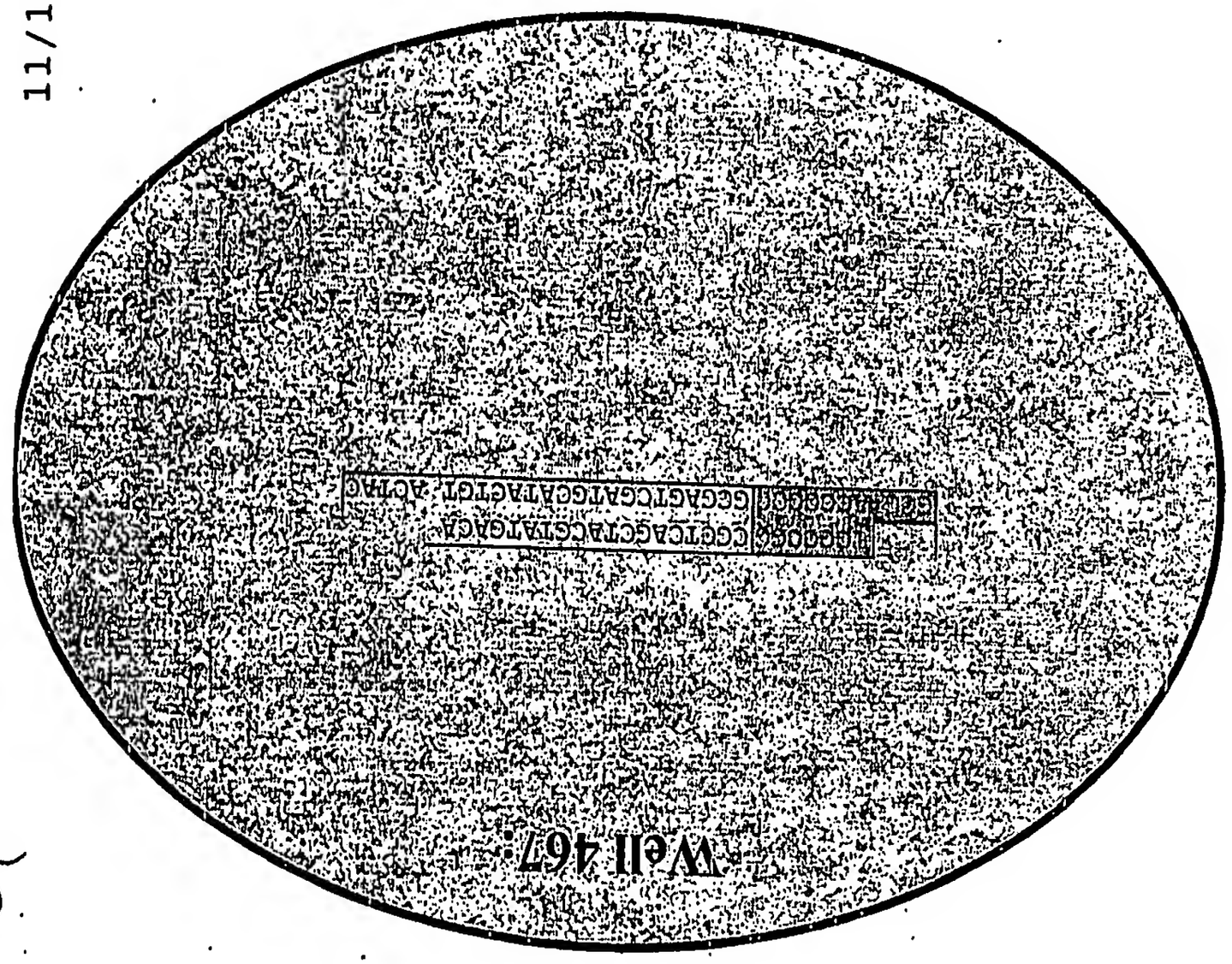

Cut with SfiI
Wash out released
end fragments
Remove

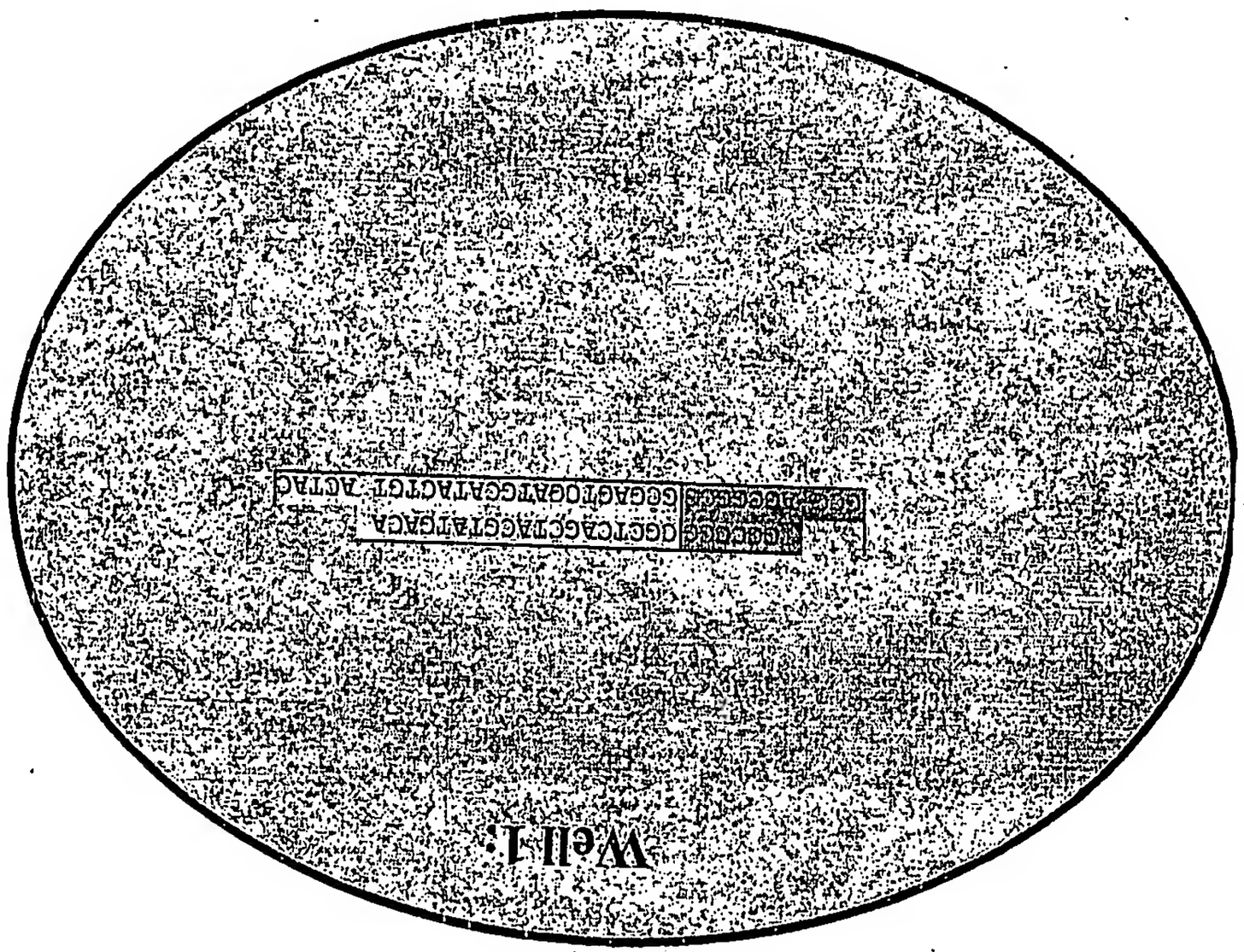

Remove
Cut with HgaI
Wash out released end fragments
Cut with SfiI
Remove

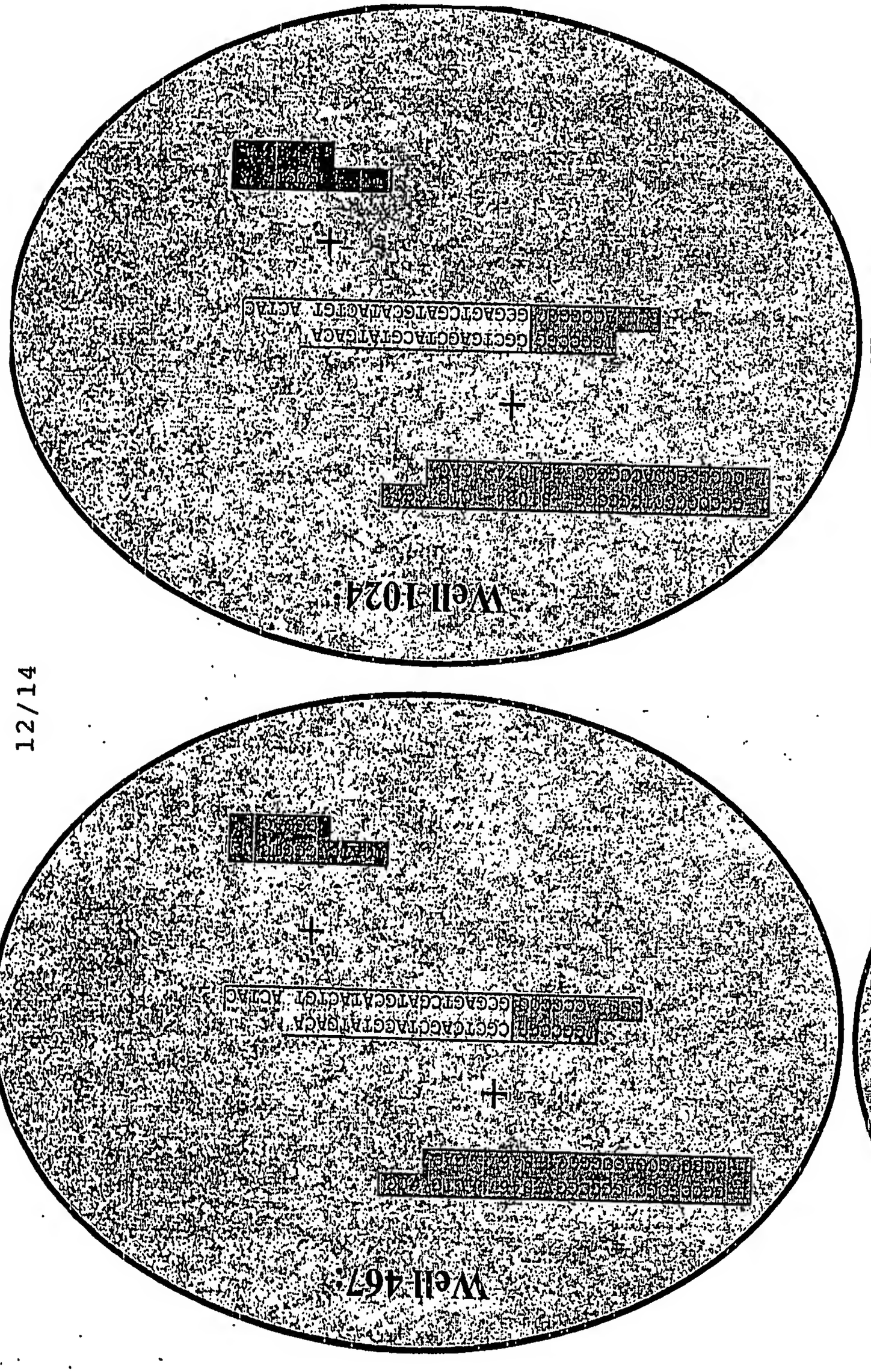

Remove
Cut with HgaI
Wash out released
end fragments
Cut with SfiI
Remove
Divide the released
fragments into 1024 wells


Well 1024:


Well 468:


Well 467:


Well 1:

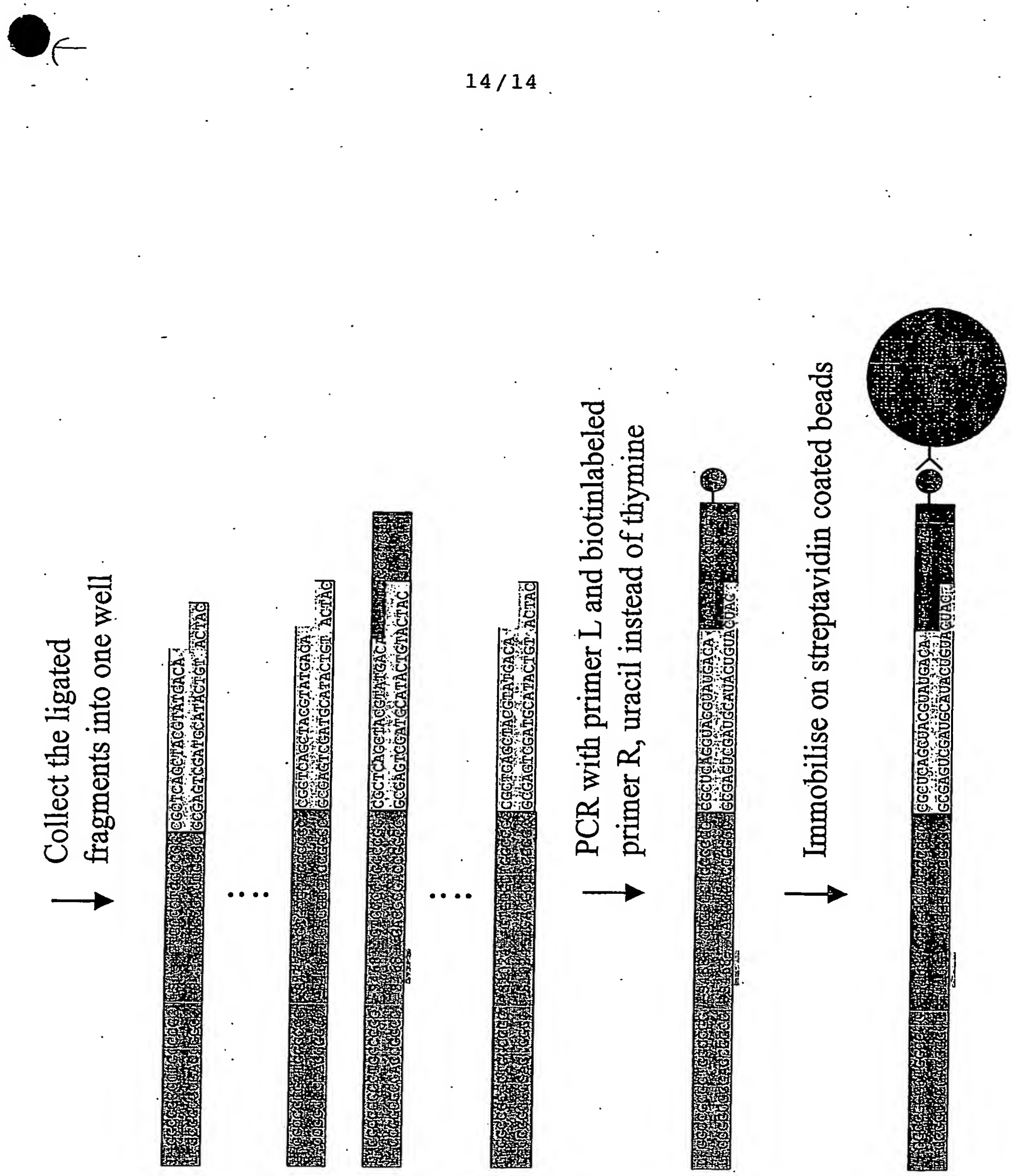

Well 1:
Well 468:
Well 467:
Well 1024:

Well 1:

Well 468:

Well 467:

Well 1024:


Collect the ligated fragments into one well
PCR with primer L and biotinlabeled primer R, uracil instead of thymine
Immobilise on streptavidin coated beads